AD_____

Award Number: W81XWH-08-1-0572

TITLE: RNAi AS A ROUTINE ROUTE TOWARD BREAST CANCER THERAPY

PRINCIPAL INVESTIGATOR: Gregory J. Hannon, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

REPORT DATE: September 2012

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2012	Annual	1 September 2011 - 31 August 2012

4. TITLE AND SUBTITLE

RNAi AS A ROUTINE ROUTE TOWARD BREAST CANCER THERAPY

5a. CONTRACT NUMBER

5b. GRANT NUMBER

W81XWH-08-1-0572

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Gregory J. Hannon, Ph.D.

E-Mail: hannon@cshl.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

During the first year of this innovator award, we made significant progress toward two of our aims. We constructed a third generation RNAi library and made that available to the breast cancer community. This resource has nearly 75,000 independent, sequence verified clones targeting ~18,000 human genes. A similar library for the mouse genome is nearing completion. We also scaled up our shRNA screening platform in preparation for lethality surveys of all suitable and available BC cell lines, including matched pairs of lines that have acquired resistance to herceptin in vitro. Relevant to our second aim, we have profiled microRNA from each of the identifiable epithelial cell types in the mouse mammary gland and are undertaking similar efforts in human. The goal is to develop microRNA sensor strategies that will permit visualization of each cell type in vivo and enable their isolation and manipulation in vitro. Finally, we showed that two microRNAs, let-7 and miR-93, can deplete tumor initiating cells from a number of basal breast cancer cell lines.

15. SUBJECT TERMS

RNAi, sequencing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	15	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

The goal of this innovator award is to continue to develop and apply RNAi-based screening methods to discover new routes toward breast cancer therapy. The project has three goals. The first is to integrate genomic and genetic information on available breast cancer cell lines to identify tumor-specific vulnerabilities and to understand genetic determinants of therapy resistance. The second is to probe the roles of breast cancer stem cells, with a particular emphasis on microRNAs. The third is to examine genomic regions that determine familial susceptibility to breast cancer using novel, focal re-sequencing methods developed in the laboratory.

Body

Fourth-generation shRNA libraries

We have developed a multiplexed validation assay for measuring shRNA potency called the 'sensor assay' (in collaboration with Steve Elledge and Scott Lowe laboratories). This assay was used to generate a large dataset of more than 250,000 measurements of shRNA efficacy (validating hairpin potency of our third-generation, human, shRNA library) from which a predictive algorithm for shRNA design, called shERWOOD, was derived. This algorithm is able to predict the results of sensor testing of shRNAs *in silico*. In addition, we tested the idea of changing the small RNA guide so that it contained a 5' U after predicting on every position of the transcriptome. That 5' residue has been shown to reside in a binding pocket of the RNAi effector complex (RISC) which favors interaction with U, but the residue is irrelevant to target recognition. Incorporating this modification into the algorithm produced even higher scores for predicted shRNAs.

Since the previous update, we have made significant progress in construction and sequence verification of our fourth generation (V4) shRNA libraries. The human library is currently comprised of 70,590 unique, sequence verified clones targeting 18,548 genes. The hairpin coverage per gene is illustrated below in the left panel. The right panel shows how many genes have at least the indicated number of hairpins at a given score (shERWOOD). For example, there are approximately 7500 genes with at least three hairpins with a score of greater than or equal to 1. Highly potent shRNAs have scores >1. Scores for shRNA designs represented in the categories of > 8 shRNAs per gene are not shown.

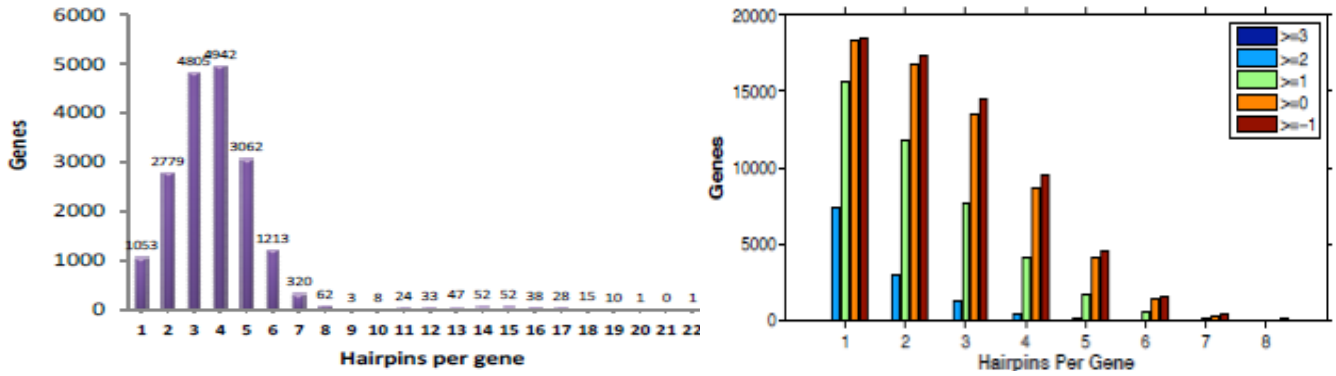

Figure 1: Human V4 shRNA library coverage by hairpins per gene (left panel) and by shERWOOD scores (right panel).

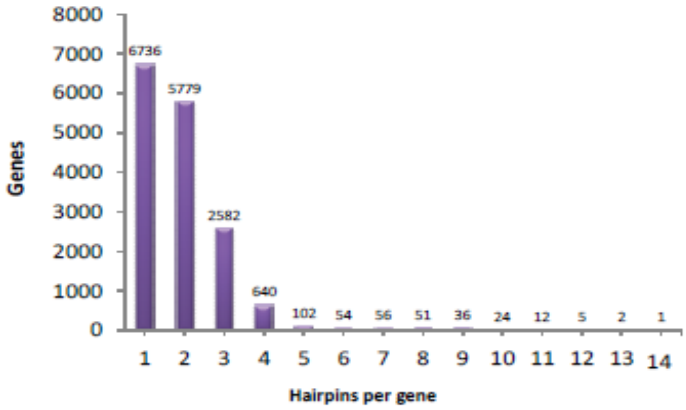

Figure 2: Mouse V4 shRNA library coverage by hairpins per gene.

In addition, we are also constructing a fourth generation mouse shRNA library and to date we have produced 31,029 sequence verified clones representing 16,079 genes.

Genome-wide RNAi screens for new therapeutic targets

Over the past year, we have continued to work towards large-scale RNAi screens *in vitro* (in collaboration with Steve Elledge's lab) using tumor-derived cell line models that are sensitive or resistant to targeted therapies (trastuzumab, lapatinib, and tamoxifen) as well as ER-positive breast cancer cell lines that are sensitive or resistant to estrogen deprivation. Our goal is to apply genome-wide, loss-of-function RNAi screens to uncover vulnerabilities of breast cancer cells in all subtypes and discover genes and pathways that modify responses to targeted therapies for *de novo* and acquired resistance.

Breast Cancer Cell Lines	Screening Conditions	Status
Her2+ treatment category		
JIMT1	No drug (straight-lethal)	Screen completed / sequencing completed
JIMT1	Lapatinib IC20	Screen completed / sequencing completed
MDA-MB-453	No drug (straight-lethal)	Screen completed / sequencing completed
MDA-MB-453	Lapatinib IC20	Screen completed / sequencing completed
MDA-MB-361	No drug (straight-lethal)	Screen completed / to be sequenced
MDA-MB-361	Lapatinib IC20	Screen completed / to be sequenced
EFM-TR	No drug (straight-lethal)	To be screened
EFM-TR	Trastuzumab (15ug/ml)	Screen completed / to be sequenced
EFM192A	No drug (straight-lethal)	To be screened

EFM-192A	Trastuzumab (15ug/ml)	To be screened
SkBr3	No drug (straight-lethal)	Screening in progress
SkBr3	Trastuzumab (15ug/ml)	Screening in progress
Sk-TR	No drug (straight-lethal)	Screening in progress
Sk-TR	Trastuzumab (15ug/ml)	Screening in progress
HCC1954	No drug (straight-lethal)	Screen completed / microarray analysis completed

ER+ treatment category

ZR75-1 Parental	+ E2	Screen completed / sequencing completed
ZR75-1 Parental	- E2	Screen completed / sequencing completed
ZR75-1Parental	- E2 / + Tamoxifen	Screen completed / to be sequenced
ZR75-1-EDR	+ E2	Screen completed / sequencing completed
ZR75-1-EDR	- E2	Screen completed / sequencing completed
ZR75-1-TAMR	+ E2	Screen completed / to be sequenced
ZR75-1-TAMR	- E2 / + Tamoxifen	Screen completed / to be sequenced
MCF7 Parental	+ E2	Screen completed / to be sequenced
MCF7 Parental	- E2	Screen completed / to be sequenced
MCF7 Parental	- E2 / + Tamoxifen	Screen completed / to be sequenced
MCF7 -EDR	+ E2	Screen completed / to be sequenced
MCF7 –EDR	- E2	Screen completed / to be sequenced
MCF7-TA MR	+ E2	Screen completed / to be sequenced
MCF7-TAMR	- E2 / + Tamoxifen	Screen completed / to be sequenced
T47D	No drug (straight-lethal)	Screen completed/microarray analysis completed

TN/Basal treatment category

Hs578T	No drug (straight-lethal)	Screen completed / to be sequenced
MDAMB231	No drug (straight-lethal)	Screen completed / to be sequenced
MDAMB468	No drug (straight-lethal)	Screen completed / to be sequenced
MDAMB436	No drug (straight-lethal)	Screen completed / microarray analysis completed
HCC1143	No drug (straight-lethal)	Screen completed / microarray analysis completed
HCC1937	No drug (straight-lethal)	Screen completed / microarray analysis completed
SUM149	No drug (straight-lethal)	Screen completed / microarray analysis completed

SUM1315	No drug (straight-lethal)	Screen completed / microarray analysis completed

Normal cells

HMEC	No drug (straight-lethal)	Screen completed/microarray analysis completed

We have completed 32 genome-wide RNAi screens in triplicate using our second generation (75,905 shRNAs targeting 19,011 genes) and third generation (74,304 shRNAs and targeting over 19,000 genes) shRNA libraries. In total there are more than 300 samples (just from the Hannon Lab) to deconvolute. To date, we have deconvoluted (by Illumina sequencing or microarray analysis) and analyzed data for 16 of the 32 screens (eight screens from the Hannon group and eight screens from the Elledge group). All remaining samples will be sequenced.

Her2-positive treatment subgroup

The HER2 (*erbB2/neu*) oncogene encodes a receptor tyrosine kinase that is amplified or overexpressed in 20-25% of human breast cancers. Patients whose breast cancers contain this alteration have an aggressive form of the disease with significantly shortened disease free and overall survivals. Traztuzumab is a recombinant humanized antibody that targets the extracellular domain of HER2. This targeted agent is efficacious in both early and metastatic HER2-positive breast cancers. In addition, the dual kinase inhibitor (HER1/EGFR and HER2) lapatinib has been shown to have clinical benefit against the metastatic HER2-postive subtype. However, not all patients whose tumors contain the HER2 alteration respond to trastuzumab or lapatinib. Less than 35% of patients with the metastatic disease respond to trastuzumab as a single agent, and more than 50% benefit from combined trastuzumab and chemotherapy. Although lapatinib has demonstrated efficacy in patients who have resistance to trastuzumab, *de novo* and acquired resistance limit its clinical potential.

We have completed all the screens from the *de novo* lapatinib resistant cell line models (JIMT1, MDA-MB-453, and MDA-MB-361). Samples for JIMT1 and MDA-MB-453 have been sequenced and MDA-MB-361 samples are in the sequencing queue.

We have analyzed the genome-wide RNAi screen data of JIMT1 (no drug) and MDA-MB-453 (no drug) for common genes that are predicted to be essential or proliferative pathways of *de novo* lapatinib resistance. This common gene list was filtered against putative essential genes for the ER-positive cell line ZR75-1 to remove those genes that might also be essential for ER-positive breast cancer cells. This analysis produced a list of candidate genes that is specific for Her2-driven cancer cells. Molecular pathways that are enriched for this set of genes include the cell cycle, protein ubiquitination, proteasome, organelle biogenesis and organization, and others.

Among the candidate genes is LGR5/GPR49, a cell surface marker involved in self-renewal in normal and cancer cells (e.g. colon cancer). Also of note is TOP1 (topoisomerase I), one of the genes that is predicted to be essential for JIMT1 and MDA-MB-453 cells to survive. We will validate a selected list of targets including TOP1 (using both RNAi and small molecule inhibition with irinotecan) and LGR5 for dependency of *de novo* lapatinib resistant cells for survival/proliferation.

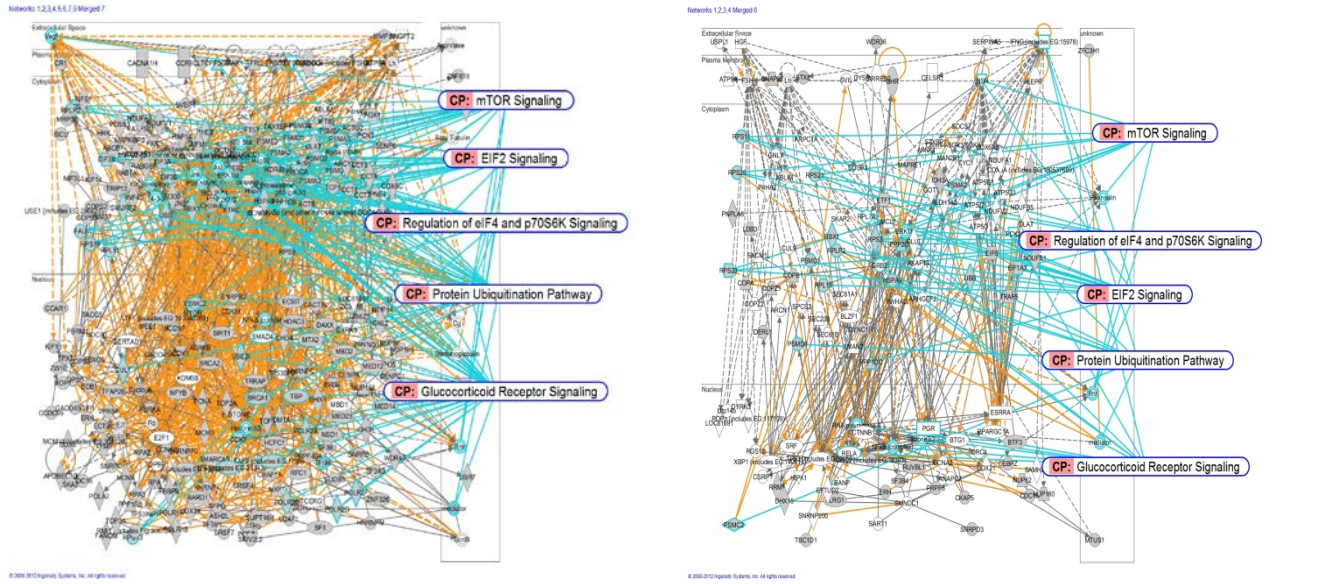

Figure 3: Interaction network of lapatinib-sensitizing genes for MDA-MB-453 (left panel) and JIMT1 (right panel) cell lines (Ingenuity).

We have also analyzed the data to inform us of potential modifiers of lapatinib resistance, particularly genes that could be targeted to sensitize *de novo* lapatinib resistant cells to the drug. Molecular pathway enrichment analysis of genes common to both JIMT1 and MDA-MB-453 suggests that several molecular complexes could be targeted to sensitize lapatinib resistant cells to the drug, including the APC/C (anaphase promoting complex/cyclosome), proteasome, and coatamer complexes. Other highly enriched cellular pathways include mTOR, EIF2, EIF4/p70S6 kinase, glucocorticoid receptor, and protein ubiquitination (Figure 3).

Validation will be carried out on a panel of *de novo* lapatinib resistant cell lines (including JIMT1 and MDA-MB-453), lapatinib sensitive lines, and normal (immortalized) human epithelial cells (HMEC) *in vitro*. Promising candidates will be further tested for their ability to sensitize lapatinib resistance *in vivo*.

ER-positive treatment subgroup

Estrogen receptor is the major driver in a majority of human breast cancers and it is expressed in 75% of breast cancers overall. Antihormone therapy is used to treat ER-positive breast cancer by either antagonizing the activity of the estrogen receptor to prevent estrogen from promoting growth of breast cancer cells (using selective estrogen receptor modulators or SERMs, e.g. tamoxifen) for premenopausal women, or by depriving cancer cells of estrogen (aromatase inhibitors) for postmenopausal women. Acquired antihormone resistance occurs when ER positive cancer cells no longer respond to this treatment paradigm.

To understand the molecular mechanisms of acquired estrogen derivation resistance (EDR) and to find target genes that would be essential for acquired EDR cells to survive, we performed whole genome RNAi screens on ZR75-1 and MCF7 tumor cell lines and their EDR derivatives. Samples for MCF7 and MCF7-EDR screens

are being prepared for sequencing. Data analysis for the ZR75-1-parental and ZR75-1-EDR screens produced a set of candidates that are proliferative or essential genes for ZR75-1-EDR but not for ZR75-1-parental cells. Figure 4 shows some of the highly enriched pathways represented by this gene set.

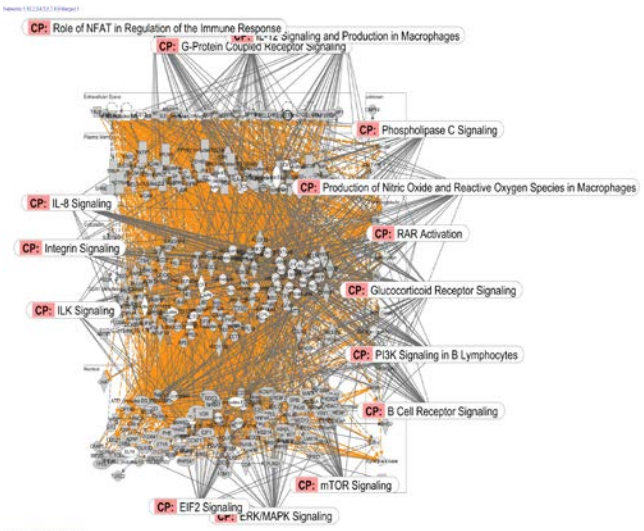

Figure 4: Interaction network of genes essential for ZR75-1-EDR cells.

Although anti-estrogen therapy is a treatment in which growth of ER-positive tumors can be attenuated, paradoxically, high-dose estrogen has also been demonstrated to cause tumor regression in postmenopausal women whose breast cancers belong to this subtype. The duration of post-menopausal period is a crucial factor affecting the success of high-dose estrogen therapy. For example, women who experienced menopause for less than one year before therapy did not respond to the synthetic estrogen (DES), while 22% of women who had reached menopause more than ten years ago responded.

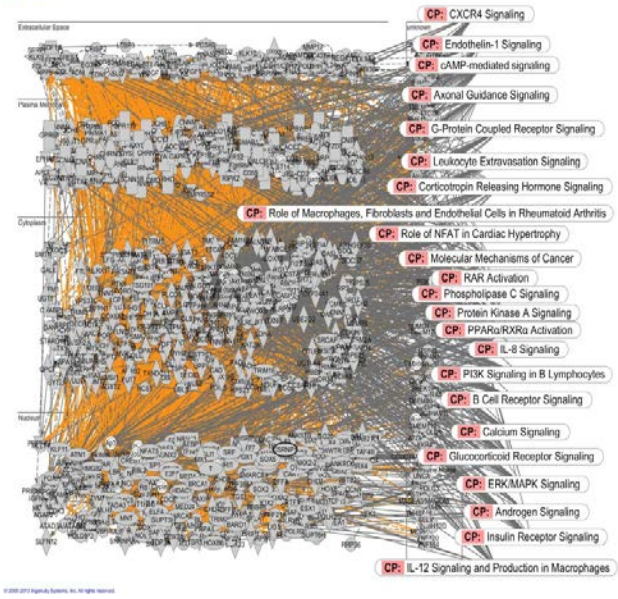

Figure 5: Interaction network of sensitizers of E2 in ZR-75-1-EDR cell line.

To investigate the mechanism of estrogen-additive therapy, cellular models of estrogen-deprivation resistance (ZR-75-1 and MCF7 cells) were used in RNAi screens to uncover modifiers of estrogen (E2) response. We have completed deconvolution (Illumina sequencing) and analysis of data for the ZR-75-1-EDR cell line and found 467 genes with more than two shRNAs that scored (37 genes with three shRNAs and three genes with four shRNAs) as sensitizers of E2. Interestingly, two of the three genes that scored with four different shRNAs are involved in fatty acid metabolism. Figure 5 represents the gene interaction network and the highly enriched canonical pathways represented by genes that demonstrated more than a two-fold depletion in the presence of E2.

We will validate a list of targets with multiple shRNA hits, including genes in pathways illustrated in Figure 5 in both ZR75-1-EDR and MCF7-EDR cells, as well as a panel of Her2-amplified and basal-like breast tumor derived cell lines and normal

mammary cells. Candidates that demonstrate the most promise will be further validated for their potential *in vivo*.

Epigenetic characterization of the mammary epithelial lineage

During the past few years we have developed a novel method to improve the isolation of mouse mammary gland stem cells (as indicated in the 2011 grant report), and have fully characterized the DNA methylation status and gene expression pattern of all mouse mammary gland cells of nulliparous mice (virgin mice), parous mice (mice with two or three sets of full pregnancy), and hormone treated mice (mice treated with three cycles of estrogen/progesterone slow release pellets). We also optimized isolation and sorting of human mammary gland cells; however due to our inability to collect the donor's pregnancy status we did not proceed with whole genome methylation profiling.

In a successful collaboration with the computational biologist Andrew Smith at University of Southern California, we have characterized methylation of all mouse mammary cell types of nulliparous and parous glands. We have spotted numerous changes in DNA methylation acquired post-pregnancy, with the most prominent modifications associated with STAT binding sites. We are currently preparing this data for publication. Outlined below is a summary of progress during 2012.

The mouse mammary stem cells (MaSCs) can be enriched at the ratio of 1:64 using a specific combination of cell surface markers. In order to improve the isolation of MaSCs we assessed the feasibility of using a transgenic mouse model of long-term label-retaining cells (K5tTA-H2b-GFP), given that a slower division rate is an accepted characteristic of adult stem cells. In this particular model, treatment with doxycycline (DOX) will cease expression of transgenic H2b-GFP, and as cells divide unlabeled H2b replaces the H2b-GFP; therefore the more slowly dividing cells will retain GFP expression for an extended period.

Histological sections revealed the presence of several GFP^+ cells located within structures resembling the mammary gland ductal epithelium, whereas treatment of H2b-GFP mice with DOX dramatically reduced the number of cells expressing GFP and those that remained GFP^+ were located at the tips of the terminal end bud (TEB) areas currently believed to contain MaSCs (Figure 6A). We next tested the ability of DOX-treated, FACS-sorted H2b-GFP- and H2b-GFP+ cells to reconstitute a new mammary gland and concluded that H2b-GFP+ cells have a five-fold greater MaSC frequency than H2b-GFP- cells (Figure 6B). The MaSC enrichment enabled by H2b-GFP+ cells was further improved by testing the transplantation activity of cells expressing cell surface markers selected according to their levels of mRNA expression (Figure 6C). One of the tested cell surface markers, Cd1d, increased MaSC enrichment by nearly ten-fold (Figure 6D), and therefore represents a novel strategy for the isolation and purification of mouse MaSCs.

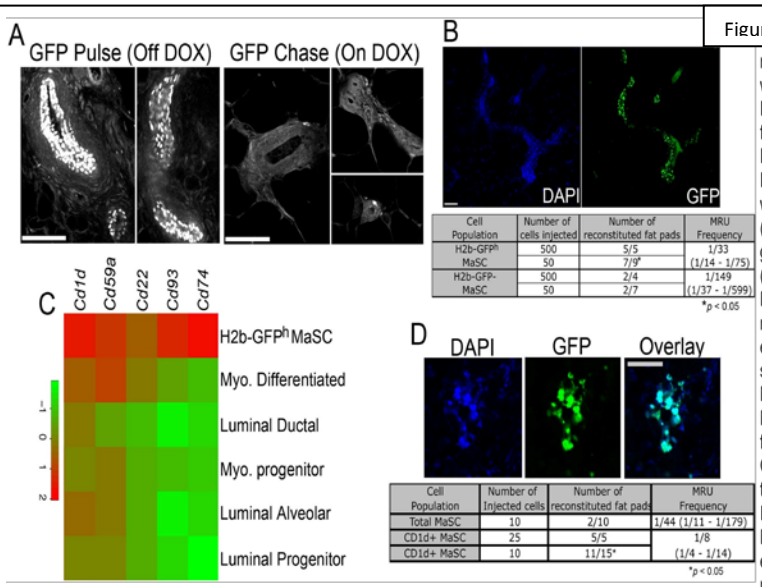

Figure 6 **Isolation of mouse MaSCs:** (A) Tissue histology of mammary glands from transgenic mice off and on DOX diet were harvested and imaged with a two-photon microscopy; (B) Histological of H2b-GFP^h MaSCs outgrowths. Cleared fat pads from pre-pubescent female mice were injected with either total H2b-GFP$^-$ negative MaSCs (CD24$^+$CD29^{h}GFP$^-$ cells) or with H2b-GFP^h MaSCs (CD24$^+$CD29^{h}GFP^{h}cells), harvested 12 weeks after transplantation and imaged on a Zeiss 710 LSM (Zeiss) confocal microscope. Images display outgrowths of gland injected with H2b-GFP^h MaSCs; Table displays MaSC (MRU) frequency according estimation by ELDA software. Minimum of 25 outgrowths are required to be considered a reconstituted gland. (C) Heatmap of cell surface markers expression across all mammary gland cell types profiled. Those shown are the most abundantly expressed within the H2b-GFP^h MaSCs. (D) Histological analysis of mammary gland CD1d MaSCs outgrowths. Cleared fat pads from pre-pubescent female mice were injected with (D) either Total MaSCs or with CD1d MaSCs. Glands were harvested 12 weeks after transplantation and endogenous GFP signal was imaged. Images display outgrowths from a gland injected with CD1d MaSCs cells; Table displays MaSC (MRU) frequency according estimation by ELDA software. Minimum of 25 outgrowths are required to be considered a reconstituted gland.

Having established that H2b-GFP^h (GFP^h=GFP+) MaSCs have mammary gland reconstitution properties, we endeavored to characterize their global mammary gland DNA methylation patterns. Using a combination of cell surface markers, six distinct cell types were isolated via FACS to a purity of >90%: H2b-GFP MaSCs (Lin$^-$ CD24$^+$CD29^{h}H2b-GFP^{h}CD61$^-$), myoepithelial progenitor cells (Lin-CD24$^+$CD29^{h}H2b-GFP$^{-/l}$CD61$^+$), myoepithelial differentiated cells (Lin$^-$CD24$^+$CD29^{h}H2b-GFP$^-$CD61$^-$), luminal progenitor cells (Lin$^-$CD24^{h}CD29$^+$CD61$^+$CD133$^-$), luminal ductal cells (Lin$^-$ CD24^{h}CD29$^+$CD61$^-$ CD133$^+$), and luminal alveolar cells (Lin$^-$ CD24^{h}CD29$^+$CD61$^-$CD133) (Figure 7).

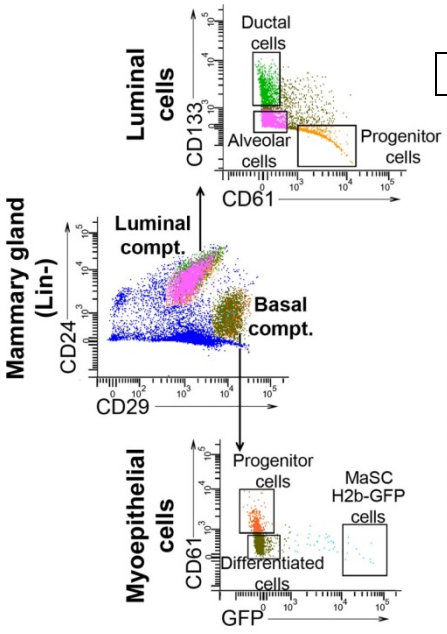

Figure 7 **Mammary gland cells sorting strategy:** We used a combination of 4 cell surface markers (ref), in addition to H2b-GFP expression, to segregate the lineage depleted mammary gland cells into 6 distinct cell types: H2b-GFP^h MaSCs (Lin$^-$ CD24$^+$CD29^{h}H2b-GFP^{h}CD61$^-$), myoepithelial progenitors cells (Lin-CD24$^+$CD29^{h}H2b-GFP$^-$CD61$^+$), myoepithelial differentiated cells (Lin$^-$CD24$^+$CD29^{h}H2b-GFP$^-$CD61$^-$), luminal progenitor cells (Lin$^-$ CD24^{h}CD29^{l}CD61$^+$CD133$^-$), luminal ductal cells (Lin$^-$CD24^{h}CD29^{l}CD61$^-$ CD133$^+$) and luminal alveolar cells (Lin$^-$CD24^{h}CD29^{l}CD61$^-$ CD133$^-$). For each library two biological replicates were analyzed.

In all sequenced samples, we achieved an optimal genome read coverage (with a mean of nine-fold), enabling us to interrogate the status of the majority of CpG sites in the genome. Hierarchical clustering of the methylation levels on promoter-associated CpGs effectively separated the six cell types into two major branches (Figure 8A). The same compartment clustering was demonstrated after pair-wise comparision among all different cell types

of the levels of methylation of Differentiated Hypomethylated Regions (DMRs) (Figure 8B). The notion that mammary gland cells are segregated into two compartments was first suggested based on gene expression analysis of murine and human cells.

We next defined luminal-differentiated DMRs (luminal alveolar and luminal ductal cell types) (Figure 8C, far left panel) and myoepithelial differentiated DMRs (Figure 8C, left panel) and plotted the levels of DNA methylation for H2b-GFP+ cells for the same regions. Patterns of DNA methylation of H2b-GFP+ cells greatly overlapped with those of basal differentiated cells, supporting the idea that a MaSC-enriched population is biased towards the basal compartment. Conversely, analysis of H2b-GFP DNA methylation levels in luminal progenitor DMRs (Figure 8C, right panel) and in myoepithelial progenitors DMRs (Figure 8C, far right panel) revealed a more intermediate methylation status, but still basally-biased, at regions where luminal progenitors and basal progenitors showed opposing methylation patterns. This observation could suggest that differentiation from a more stem-like cell type to a more lineage-committed cell type involves both acquisition and loss of DNA methylation. Regulation of epigenetic mechanisms at the mammary gland stem cell level is important in the control of self-renewal and differentiation, since the default condensed methylation levels in stem cells accommodate changes in DNA methylation that would dictate lineage specificity, a hypothesis experimentally supported in variety of tissues.

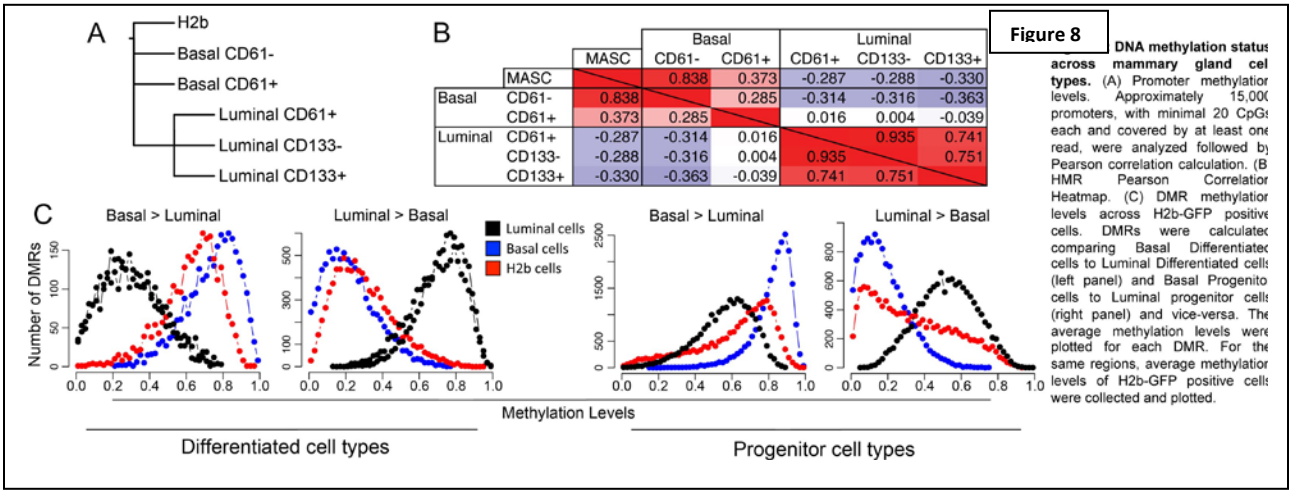

Figure 8

DNA methylation status across mammary gland cel types. (A) Promoter methylatior levels. Approximately 15,00(promoters, with minimal 20 CpGs each and covered by at least one read, were analyzed followed by Pearson correlation calculation. (B HMR Pearson Correlatior Heatmap. (C) DMR methylatior levels across H2b-GFP positive cells. DMRs were calculatec comparing Basal Differentiatec cells to Luminal Differentiated cells (left panel) and Basal Progenito cells to Luminal progenitor cells (right panel) and vice-versa. The average methylation levels were plotted for each DMR. For the same regions, average methylatior levels of H2b-GFP positive cells were collected and plotted.

In order to understand how DNA methylation orchestrates mammary gland cell differentiation or lineage specification we carried out RNAseq for each one of the cell types and computed their RPKM values. We next defined three sets of differentially expressed genes: H2b upregulated genes, basal upregulated genes (all genes commonly upregulated in H2b-GFP cells, myoepithelial progenitor cells and myoepithelial differentiated cells), and luminal upregulated genes (all genes commonly upregulated in luminal progenitor cells, luminal alveolar cells and luminal ductal cells). Each set contained approximately 50 genes (Figure 9A). We next collected data regarding methylation levels surrounding the transcription start site (TSS) of genes upregulated in each one of these gene pools (Figure 9B). In all six mammary gland cell types, genes differentially expressed in H2b cells displayed unchanged DNA methylation levels upstream of the TSS and slightly lower levels downstream of the TSS

relative to genes that were not differentially expressed. The landscape of methylation levels across the TSS of upregulated genes displayed a greater degree of differential methylation 1-2kb downstream of the TSS. Collectively, our results contributed to the elaboration of the first mouse mammary methylome and provided important insights about the dynamics of DNA methylation across a spectrum of mammary gland cell types. We are currently analyzing DNA methylation libraries from CD1d-isolated MaSCs to further improve our knowledge of DNA methylation dynamics of mammary gland cells.

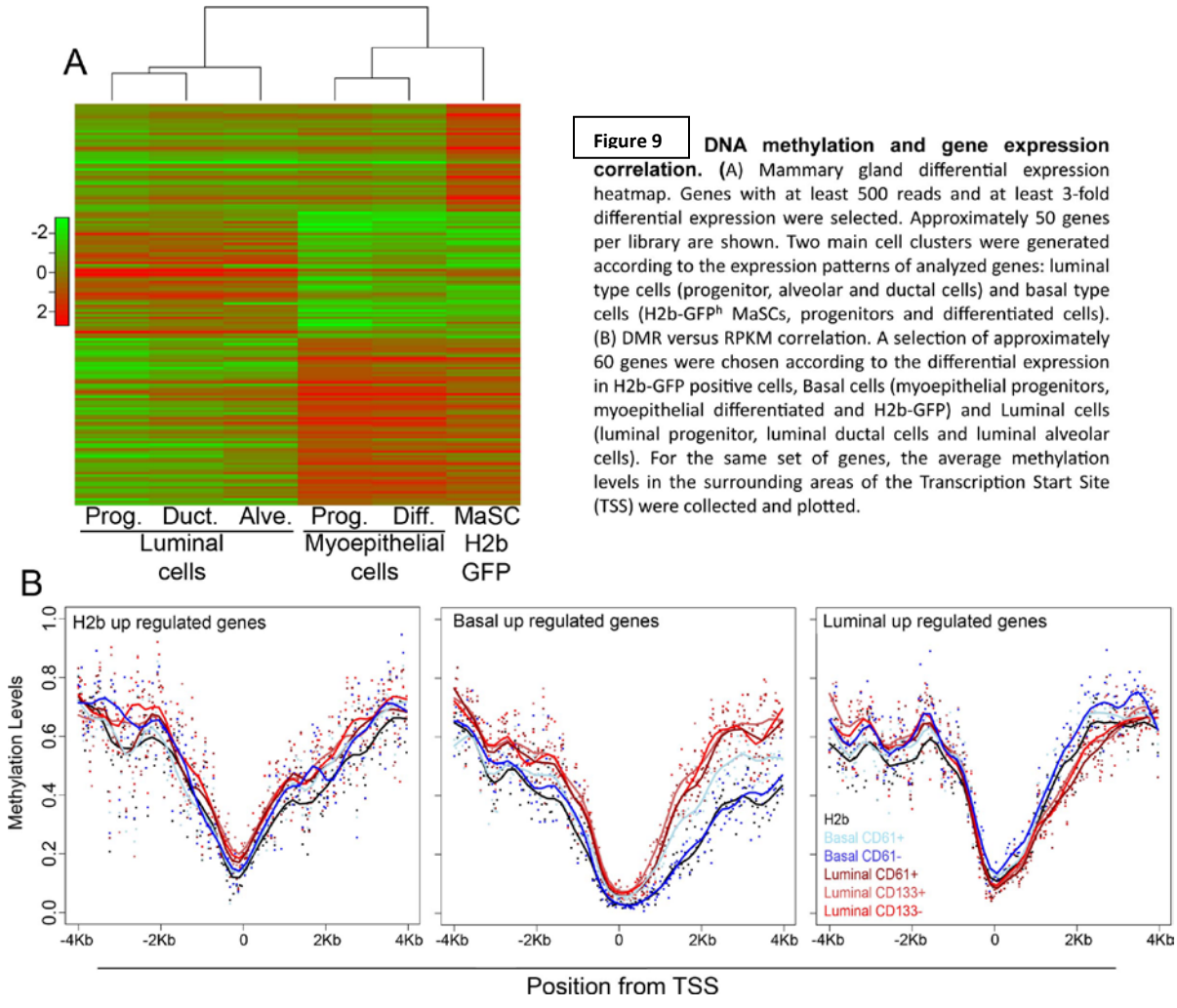

Figure 9 **DNA methylation and gene expression correlation.** (A) Mammary gland differential expression heatmap. Genes with at least 500 reads and at least 3-fold differential expression were selected. Approximately 50 genes per library are shown. Two main cell clusters were generated according to the expression patterns of analyzed genes: luminal type cells (progenitor, alveolar and ductal cells) and basal type cells (H2b-GFP^h MaSCs, progenitors and differentiated cells). (B) DMR versus RPKM correlation. A selection of approximately 60 genes were chosen according to the differential expression in H2b-GFP positive cells, Basal cells (myoepithelial progenitors, myoepithelial differentiated and H2b-GFP) and Luminal cells (luminal progenitor, luminal ductal cells and luminal alveolar cells). For the same set of genes, the average methylation levels in the surrounding areas of the Transcription Start Site (TSS) were collected and plotted.

Having documented the DNA methylation signature of all mammary cells from nulliparous (virgin) mammary gland, we next generated parous mammary methylome libraries using the same cell sorting strategy described above. Female mice were allowed two full pregnancy cycles, including birth, nursing and full involution (two months). Due to increased cell division rates, no H2b-GFP+ cells were present in the glands of parous mice. We are currently preparing DNA methylation libraries from CD1d-isolated MaSCs to investigate the effects of pregnancy in the MaSC

compartment. Genomic coverage for the parous libraries resembles that achieved for the nulliparous methylome (approximately 9-fold coverage).

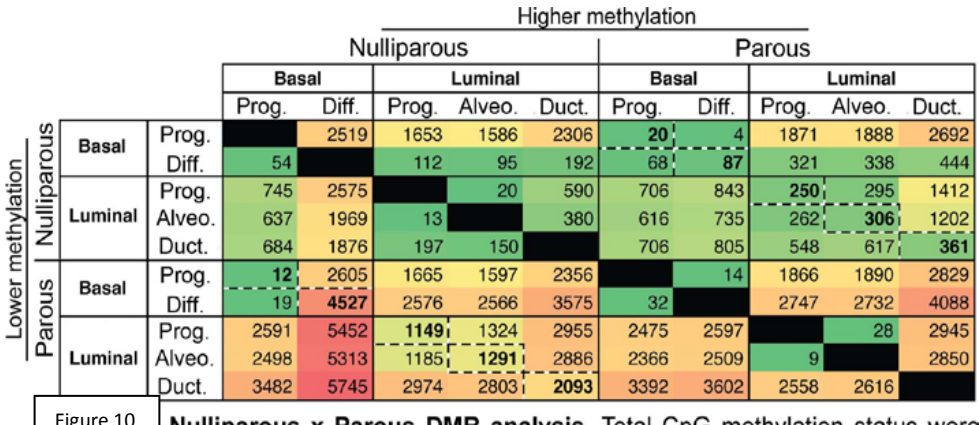

Figure 10 **Nulliparous x Parous DMR analysis**. Total CpG methylation status were analyzed and hypomethylated regions (HMRs) calculated. Nulliparous or parous HMRs were analyzed according to Parous or nulliparous methylation status, and regions with differential methylated (DMRs, at least 10 significantly differing CpG per DMR) were plotted.

In order to map DMRs we analyzed both libraries in a bidirectional pair-wise fashion, by comparing each cell to its corresponding cell type before and after pregnancy (Figure 10). The amount of nulliparous DMRs (lower methylation levels before pregnancy) was substantially smaller (dashed line, upper right side) than the number of parous DMRs (lower methylation levels after pregnancy), suggesting a dramatic loss of methylation by most cell types post-pregnancy (dashed line, lower left side). The loss of methylated sites after pregnancy could translate into changes in gene expression, an observation that was previously suggested to be the case for a small subset of genes. We are currently comparing RNAseq libraries of all mouse mammary cell types before and after pregnancy to more precisely identify the changes in gene expression patterns.

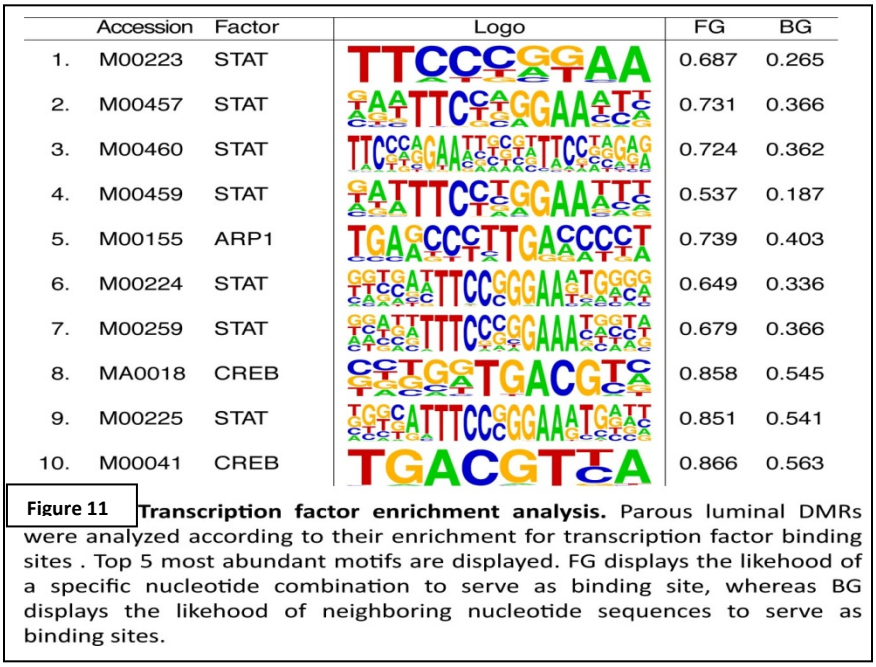

Figure 11 **Transcription factor enrichment analysis.** Parous luminal DMRs were analyzed according to their enrichment for transcription factor binding sites . Top 5 most abundant motifs are displayed. FG displays the likehood of a specific nucleotide combination to serve as binding site, whereas BG displays the likehood of neighboring nucleotide sequences to serve as binding sites.

The luminal compartment, exhibited the most DNA methylation changes after pregnancy. The extent of these differences was reflected in the number of acquired hypomethylated sites (DMRs) but was most importantly also correlated with DNA methylation loss. Among all luminal cell types (progenitor cells, alveolar cells and ductal cells), a great portion of shared DMRs occurred nearby binding sites recognized by the STAT

transcription factor (Figure 11), which might suggest a role for this family of proteins during pregnancy, lactation and involution. Interestingly, the STAT-associated DMRs were further enriched for a class of genes with known roles in apoptosis and potential antitumor activity.

STAT transcription factors have been previously suggested to play important roles in mammary cells. STAT5a and STAT5b have been implicated in the transcriptional activation of milk protein in response to progesterone levels, although STAT5a and STAT5b protein levels only slightly increased during pregnancy and lactation. Lack of STAT5a expression resulted in decreased lobuloalveolar development during normal mammopoiesis and blocked milk production in the first pregnancy, although ductal density and milk production resumed at the onset of a second pregnancy. Conversely, overexpression of full-length STAT5a not only induced lobuloalveolar development but also delayed involution, whereas overexpression of a c-terminally truncated form of STAT5a accelerated apoptosis during involution. Further understanding of how STAT transcription factors regulate gene expression in mammary cells, including how this regulation is susceptible to changes during pregnancy could provide a clear foundation for evaluating the role of STATs in pregnancy-induced breast cancer protection.

Reportable Outcomes

1. A fourth generation human shRNA library comprised of 70,590 shRNAs targeting 18,548 genes.
2. A fourth generation mouse shRNA library comprised of 31,029 shRNAs targeting 16,079 genes.

Conclusions

We have made significant progress over the past year and we will continue to make progress toward the major goals of this application. For 2012, we have begun to deconvolute our genome-wide RNAi studies, and have identified candidate genes/pathways that are essential for *de novo* lapatinib resistance as well as genes that can potentially sensitize these resistant cells to lapatinib for new combination therapy. Furthermore, we have initiated a study to understand resistance to estrogen-deprivation and we will continue to deconvolute our remaining genome-wide screens in the coming year. This past year, some of the most important strides have been made in the understanding of mammary epithelial biology through our epigenetic characterization of mammary epithelial cell lineages.